A statistical criterion for species and genera among the bacteria*

C.-E. A. Winslow

The existence of an almost infinite number of minute variations in many groups of bacteria has so far almost nullified any attempt at a natural classification. Isolated bacterial cells, lying free in a liquid medium, are exposed in a high degree to the direct effect of the environment; and the extent to which virulence may be exalted or attenuated shows how easily they respond to such influences. The absence of the sex process probably magnifies any tendencies to variability. Amphimixis may well serve, as Weissmann holds, to promote minor variations; but on the other hand it is reasonable to believe that it checks more extreme deviations from the mean. Among asexual organisms, every variation that arises is preserved intact until acted on by some modifying cause. Finally, the rapid multiplication of the bacteria offers an exceptional opportunity for the action of selective influences. The immense number of generations which succeed each other in a short space of time makes boundary lines as shifting as they would become among the higher plants if a dozen geological periods were considered all at once.

In certain groups of the bacteria, among the Coccaceae for example, and in the group of *Bacillus Diphtheriae*, the inconstancy of characters is specially marked and it is possible by appropriate laboratory experiments to modify profoundly such important metabolic properties as acid-production and chromogenesis. Other groups, like that of the aerobic spore-formers, appear to be in relatively stable equilibrium. Even where the characters of any given strain are fairly constant, classification is often complicated by the fact that an almost infinite series of minutely differing strains may be found, each apparently constant under existing conditions but together forming a continuous series connecting widely separated types.

* Read before Section G of the American Association for the Advancement of Science at the Baltimore Meeting, December 31, 1908.

The result of this condition of affairs is that systematic bacteriology has almost fallen into desuetude. After a brave beginning in the way of describing bacterial species the conviction gradually gained ground that of the making of species there is no end. Aside from the forms which are recognized by their association with well-defined diseases, there are hardly a score of bacteria that can be distinguished with any certainty among the hundreds which have been already described. Time and industry alone would be needed to produce as many thousands as there are now hundreds of paper species; but the condition of affairs would only be still more hopeless.

Refuge from the multitude of varieties has been sought, on the one hand, by ignoring all minor differences and massing the bacteria into a few main groups; and, on the other hand, by a frankly arbitrary schematic arrangement of cultures according to their plus-or-minus reaction in a few standard media, decimal numbers being used for the description of each combination of possible reactions. Neither of these expedients is very satisfactory. The first ignores real differences of much practical as well as theoretical importance. The second method, while convenient for the cataloguing of descriptions, only obscures the natural phylogenetic relationships of bacterial types. Organic individuals have not been created in symmetrical groups, differing in the presence or absence of character *A*, each group including two sub-groups, respectively possessing and lacking character *B*. They have been developed along irregular and complex lines, and are now related in family groupings of infinitely various kinds and degrees.

There is one method, and apparently only one, which promises to answer the puzzling question of bacterial relationships. This is the statistical method. If attention be given, not to the individual, but to the group, and if certain characters are quantitatively measured in a considerable series of cultures and statistically analyzed, order begins to emerge from chaos. Almost every degree of acidity or gelatin liquefaction may be found in some individuals; but the greatest number of strains group themselves about well-defined centers. The curve of frequency plotted for any character shows certain modes which are the types about which the bacteria as a group are varying. Furthermore, if sev-

eral distinct characters are compared, it will frequently appear that on the average certain properties are correlated with each other and generally vary together. As mountain tops occur in ranges, so the peaks on the curve of frequency for any single bacterial character are grouped together, by the study of other characters, into larger complexes. Major and minor groups may rightly be considered as constituting genera and species among the micro-organisms. They are not separated by boundary lines but are fixed rather by the centers about which they group themselves. They are genera and species, characterized, not by the description of individual strains, but by the frequency of occurrence of a considerable series of such individuals.

This method of statistical classification is by no means novel. Ever since Quételet and Galton began to apply the statistical method to the solution of biological problems it has been recognized that species and varieties could be more or less satisfactorily fixed by a study of frequency of distribution. Davenport and Blankinship (Science II. **7**: 685) even went so far as to suggest a definite numerical criterion of a species, founded on this basis. According to their suggestion, if a bimodal curve is obtained, the ratio of the ordinate of the lesser mode to the minimum ordinate in the depression between the modes, may be considered as an index of the isolation between the two races. If this ratio is over two, the modes represent species; if under two, varieties. The numerical criterion fixed by these authors may be criticized as somewhat arbitrary. The general principle is sound, however, and has yielded good results in many fields. Notably in the study of the races of man the statistical method has proved of the highest service. Curiously enough, among the micro-organisms, where some method of this sort is most urgently needed, statistical study has until very recently been wholly lacking.

At the Massachusetts Institute of Technology we have been for some years working on the classification of the family Coccaceae, one of the most difficult groups of the bacteria, by this general method. We collected five hundred different strains of cocci from various sources, earth, water, air, and the normal and diseased human body. After some considerable preliminary study eleven characters were fixed upon as having probable systematic importance

and as adapted to definite quantitative measurement. When the results obtained were analyzed from the statistical standpoint, groups and subgroups of various grades were obviously apparent; and it was possible to draw up on this basis a general outline of a classification for the group, which appears to correspond to true phylogenetic relationships (Winslow, C.-E. A., & Winslow, A. R., Systematic Relationships of the Coccaceae; New York, 1908). In the first place, two major divisions or subfamilies were distinguished, one saprophytic and the other parasitic. The cocci from the human body generally occurred in chains or small irregular groups of cells, stained by the Gram stain, formed a meager or only fair surface growth on media, produced acid in carbohydrates, and showed no pigment or a white or orange one. The cocci from earth and water, on the other hand, occurred in large cell groups or packets, decolorized by the Gram stain, grew well on solid media, failed to ferment carbohydrates and produced a red or yellow pigment. Each character was sometimes found in the group where it did not normally occur, but on the whole the average correlation was very strong. Within each subfamily several groups of a second grade of individuality were found, marked by the association of a smaller number of characters than the subfamilies, but still defined by the correlation of several independent properties. These were called genera. Finally, within the genera, species were defined by the distinct centers of variation which were manifest for single independent characters, such as the liquefaction of gelatin and the reduction of nitrates.

Curiously enough, while we were at work upon the relationships of the family of the Coccaceae as a whole, Andrewes and Horder in England were applying almost exactly the same principles to the classification of the subtypes included in the genus *Streptococcus*, a genus which is so marked in its characters that it has been one of the few groups of the Coccaceae long recognized as a result of common observation. What species should be included in the genus *Streptococcus* has been the subject of long and fruitless debate; and the "Vielheit" or the "Einheit" of the *Streptococci* seemed an insoluble problem. Andrewes and Horder, however, by studying the numerical frequency of occurrence of various forms distinguished seven type centers which they could properly

recognize as species (Lancet 1906[2]: 708). Their work greatly extended ours, as we had worked scarcely at all on the particular genus with which they dealt. The combined effect of the two investigations is to emphasize strongly the value of the statistical method in bacterial classification; and there seems no reason why the same principles should not be of similar value in the systematic study of algae and fungi and Protozoa and other simple and variable forms.

Among other things that these studies emphasized is the importance of physiological differences in bacterial classification. Systematic bacteriology has been greatly retarded by the undue emphasis laid on morphological characters. Reasoning from analogy with higher plants, many bacteriologists have refused to base generic distinctions upon anything but morphological differences. As there are very few morphological differences in the group, there has been no rational generic classification. The distinction between physiological and morphological characters is merely a superficial one. Both are presumably due to chemical modifications of protoplasm; and there is no reason to suppose that a protoplasmic property which manifests itself in the size and arrangement of parts is any more fundamental than one which manifests itself in the ability to utilize a certain food stuff. It is precisely along the lines of metabolism that the bacteria have attained their extraordinary degree of differentiation. The higher plants have developed complex structural modifications to enable them to absorb food materials of certain limited kinds and to utilize the sun's energy in building them up into protoplasm. Meanwhile the bacteria have maintained themselves by acquiring the power of assimilating simple and abundant foods of the most varied sorts. Evolution has developed gross structure in one case without altering metabolism; it has produced a diverse metabolism in the other case, without altering gross structure. There is as wide a difference in metabolism between the *Pneumococci* and the nitrifying bacteria as there is in structure between a liverwort and an oak. So-called physiological differences are quite as important in one case as so-called morphological differences in the other.

The characters which are of prime systematic importance will naturally vary according to the particular group of bacteria which

is under consideration. One family or genus may have differentiated along the line of carbohydrate fermentation; another may have divided into groups adapted to a parasitic life on particular tissues of the human body, like the *Diplococci*; in a third, obvious differences in reproduction may have arisen, as in the anaerobic spore-formers. This is one reason why the conventional bacterial classifications are inadequate. Migula (System der Bakterien, II., Jena, 1900) and Chester (A Manual of Determinative Bacteriology, New York, 1901), for example, recognize only five morphological genera among the Coccaceae and divide them all into dichotomous classes by the same arbitrary plus-and-minus tests. Such an arrangement is very far from representing the complex facts of relationship. A difference of great systematic importance in one group may be valueless in another. In each group of the bacteria as many reactions as possible must be examined with an open mind in order to determine which are of systematic importance. In many groups the most helpful properties which mark their characteristic differentiation may be biochemical reactions not yet studied at all. It was the careful examination of fermentative power in various carbohydrate media by Gordon and his colleagues that made the classification of the *Streptococci* possible and it is quite probable that a similar detailed knowledge of the saprophytic *Micrococci* will be acquired only by the application of tests not now at our disposal.

The tests chosen for systematic study of the bacteria should, as far as possible, be susceptible of definite quantitative measurement. A great deal of the earlier confusion which has surrounded the subject has been due to descriptions of colony-formation, etc., which were expressed in terms wholly dependent on the personal equation of the author. For statistical study, records should, if possible, be capable of expression on a numerical scale. Where this is not feasible, as in comparing pigments of different hues, the observations may be divided into classes, defined as closely as possible by comparison with known standards. The cultures studied should include a considerable series of the particular group studied; and it should represent a wide variety of habitats, as the bacteria are usually closely adapted to their particular environment.

When the observations of several hundred cultures are finally

obtained and tabulated, the first important thing is to determine the major groups included in the series. This may be easily done by a general study of the correlation of different characters, without any elaborate mathematical methods. In our work on the Coccaceae we constructed a correlation table for each pair of characters and noted the general coincidence of one with the other.

It must be remembered, in estimating the importance of the occurrence of correlated characters, that their common presence may be due to one of two causes. The characters may be correlated simply because ancestral forms developed them both, under the action of independent causes. Or, on the other hand, the properties in question may not be really independent but subtly bound up in the physiological balance of the organism so that a modification of one leads to a corresponding change in the other. In either case profound modifications which have altered the whole center of gravity of the organism may rightly be considered of generic or family rank.

Within the genera, specific types may be defined by variations in single or subordinate characters.

The final test as to whether two related organisms deserve recognition as species, or are only variants from a single type, must be made by an examination of the curve of frequency for the character on which their identity is supposed to rest. If there is a single center of frequency of occurrence it is fair to consider that the type is a simple one, however variable it may be. Thus Goodman (Journal of Infectious Diseases **5**: 421) has recently made a study of the *B. Diphtheriae* group, which indicates that widely differing strains of these organisms belong to a single type. If on the other hand two distinct modal points indicate two separate centers of distribution each may well receive a specific name. Intermediate cases will of course be found, in which the curve of frequency has two peaks separated only by a shallow trough. In such cases Davenport and Blankinship's criterion, to which reference has been made above, may be helpful ; although any such arbitrary standard is likely to be modified by enlarged knowledge.

It may be maintained with reason that genera and species of bacteria defined by the statistical method will not be invalidated by the discovery that the characters on which they are based may be

modified, either experimentally or under natural conditions. Among the flowering plants the study of a long series of specimens from habitats separated in space or time often tends to make the boundaries between species almost indistinguishable. If we could modify and control the environment of these higher forms through countless generations, as we do in the case of the bacteria, there can be little doubt but that we could transmute one form into another. Yet for practical convenience, and in recognition of the existing preponderance of certain types, we recognize them as constituting species. So among the bacteria we may reasonably hold that even mutable types in certain unstable groups are deserving of systematic recognition.

It is a curious fact that some of the most variable characters among the Coccaceae, chromogenesis and reaction to the Gram stain, for example, proved to be most strikingly correlated with the modifications of other powers, associated respectively with the parasitic and saprophytic habitats. Among the higher plants the modifications due to environment are comparatively superficial. Among the plastic unicellular organisms on the other hand the adaptation to parasitic life, for example, may produce profound changes, which warrant generic rank. Among the more complex plants and animals, indeed, the same conception of purely environmental types is gaining ground. It is not in connection with bacteria that Jordan and Kellogg (Evolution and Animal Life, New York, 1907), speak of "Ontogenetic species held for a number of generations true to a type simply because the environment, the extrinsic factors in the development of all the individuals in these successive generations, are the same."

The main points which I have tried to emphasize as essential to a rational natural classification of the Schizomycetes may be briefly summarized as follows:

1. In groups like the bacteria and perhaps some classes of the fungi, which have differentiated along physiological rather than morphological lines, differences in metabolism may have the same systematic importance given to gross structural differences in other groups.

2. The characters of greatest systematic importance, whether morphological or physiological, vary in each particular group of

bacteria, according to the lines followed in recent evolution. The classificatory value of a particular character must be determined for each group by a preliminary survey of its relationships.

3. The most satisfactory method of studying the systematic relations of these simple and variable forms is by obtaining quantitative measurements of a number of characteristic properties in a large series of individuals and by statistical analysis of the results.

4. Generic names may conveniently be given to the larger groups of organisms having several apparently independent properties in common, as indicated by correlation between different measurements. Generally these groups will be found associated with certain definite habitats.

5. Specific names may be reserved for the smaller groups, characterized by variations in single uncorrelated properties. Each strongly marked mode on the curve of frequency, marking a type center about which variations are occurring, may be taken as a species.

6. Genera and species defined in this manner by frequency of occurrence are not invalidated by the variability of the organisms belonging to them. Stable or unstable, they are definite realities, and their recognition makes it possible to arrange the bacteria in convenient natural groups which approximate their true phylogenetic relationships.

BIOLOGICAL LABORATORIES,
MASSACHUSETTS INSTITUTE OF TECHNOLOGY,
BOSTON, MASSACHUSETTS.